ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2020)
69, *Part* 1, *pp.* 195–218

Modelling and prediction of financial trading networks: an application to the New York Mercantile Exchange natural gas futures market

Brenda Betancourt,

University of Florida, Gainesville, USA

Abel Rodríguez,

University of California, Santa Cruz, USA

and Naomi Boyd

West Virginia University, Morgantown, USA

[Received September 2018. Final revision September 2019]

Summary. Over recent years there has been a growing interest in using financial trading networks to understand the microstructure of financial markets. Most of the methodologies that have been developed so far for this have been based on the study of descriptive summaries of the networks such as the average node degree and the clustering coefficient. In contrast, this paper develops novel statistical methods for modelling sequences of financial trading networks. Our approach uses a stochastic block model to describe the structure of the network during each period, and then links multiple time periods by using a hidden Markov model. This structure enables us to identify events that affect the structure of the market and make accurate short-term prediction of future transactions. The methodology is illustrated by using data from the New York Mercantile Exchange natural gas futures market from January 2005 to December 2008.

Keywords: Array-valued time series; Financial trading network; Hidden Markov model; Stochastic block model; Systemic risk

1. Introduction

Financial trading networks are directed graphs in which nodes correspond to traders participating in a financial market, and edges represent pairwise buy–sell transactions among them that occur within a period of time. Financial trading networks contain important information about patterns of order execution in order-driven markets where all buyers and sellers display the prices at which they wish to buy or sell a particular security. Therefore, they provide insights into aspects of market microstructure such as market frictions that influence the execution of trades, trading strategies and systemic risks that could cause the collapse of the market. For example, consider the role of financial trading networks in understanding the effect of market frictions on market microstructure. In the absence of market frictions, we could expect orders from different traders to be matched randomly. However, real trading networks often exhibit

Address for correspondence: Brenda Betancourt, Department of Statistics, University of Florida, 220 Griffin-Floyd Hall, PO Box 118545, Gainesville, FL 32611-8545, USA.
E-mail: bbetancourt@ufl.edu

 0035–9254/20/69195

features such as elevated transitivity or preferential attachment among certain groups of actors (Adamic *et al.*, 2010), which are inconsistent with random matching. In the case of open outcry markets, these features can be partially explained by sociological factors (for example, see Zaloom (2004)). Alternative explanations include the effect of different market roles (e.g. liquidity providers or takers) or trading strategies (e.g. long *versus* short strategies); see Ozsoylev *et al.* (2010) or Hatfield *et al.* (2012).

Financial trading networks also provide information that is key in the assessment of systemic risks. Analysis of the evolution of the networks over time, particularly the detection of drastic structural changes, can aid in tests of financial market stability (or fragility as it may be) by financial regulators to ensure that events such as a large trader failure do not serve to destabilize financial markets. For example, identifying the failure of a large trader and understanding how it affects the network structure will help to guide regulators through the process of unwinding their positions and may dictate whether those positions are unwound in the open market or through a transfer to a suitable counterparty (Boyd *et al.*, 2011). Here, we are interested in a model for change point identification that can be used by regulators in the analysis of the dynamics of trading networks to improve market supervision. The analysis of trading networks can also be utilized to identify important traders who play a critical role in the market (e.g. by acting as *de facto* market makers or liquidity providers). Moreover, they can also help us to identify frequent counterparties of specific traders which may aid in regulatory oversight by federal agencies and market exchanges alike. Indeed, there is evidence that price distortion and manipulation may be more likely between frequent counterparties than by one agent acting in isolation (Harris *et al.*, 1994).

The literature on the mathematical modelling of financial trading networks is limited. Theoretical approaches that explain the structure of a financial network as the outcome of a game have recently been developed (for example, see Ozsoylev *et al.* (2010) and Hatfield *et al.* (2012)), but they are of limited practical applicability. Most of the empirical work on trading networks has focused on the use of summary statistics such as degree distributions, average betweenness and clustering coefficients (Newman, 2003; Adamic *et al.*, 2010). These types of approach provide some interesting insights into market microstructure but suffer from two main drawbacks. First, the summary statistics to be monitored need to be carefully chosen to ensure that relevant features of the market are captured (for an example of this, see Section 2). Although some of the game theoretic work that was mentioned before might provide insights into which network summaries should be monitored, the choice is typically *ad hoc* and the selection is often incomplete. Second, and more importantly, approaches of this type are not helpful in predicting future interactions between traders.

In this paper we move beyond descriptive network summaries to focus on stochastic models for array-valued data that place a probability distribution on the full network. The simplest such model is the classical Erdös–Rényi model (Erdös and Rényi, 1959), which assumes that interactions between any two traders occur independently and with constant probability that is independent of the identity of the traders. This class of models, although well studied from a theoretical perspective, is too simplistic to accommodate most realistic networks. As an alternative, Frank and Strauss (1986) proposed the class of exponential random graph models (ERGMs), which are also called p^*-models. These models formalize the use of summary statistics by including them as sufficient statistics in exponential family models. A temporal version of the ERGM was introduced in Hanneke *et al.* (2010) and further developed in Cranmer and Desmarais (2011) and Snijders *et al.* (2010). The class of p_1-models, which extends generalized linear models to array-valued data, was originally proposed by Holland and Leinhardt (1981) and extended to dynamic settings in Banks and Carley (1996), Goldenberg *et al.* (2009) and

Kolacyzk (2009). Another related approach was introduced in Hoff *et al.* (2002), using the concept of latent social space models. In this class of models the probability of a link between nodes increases as they occupy closer positions in latent social space. Models that are based on latent social spaces have again been extended to dynamic settings by Sarkar and Moore (2006) and Sewell and Chen (2015), among others.

The model that is discussed in this paper extends the class of stochastic block models that were first introduced in Wang and Wong (1987) to account for time dependence. Stochastic block models rely on the concept of structural equivalence to identify groups of nodes with similar interaction patterns. We are particularly interested in identifying groups of traders with similar interaction patterns which we refer to as *trading communities* in the context of this application. Our model is designed to identify trading communities where traders do not necessarily interact with each other but interact with the same others. This is an important feature of the model motivated by financial trading networks where disassortative community behaviour is key to characterize trading patterns. Model-based stochastic block models have been developed as array-valued extensions of traditional mixture models, and dynamic versions of these models have recently been proposed. For example, Nowicki and Snijders (2001) presented a simple Bayesian model that uses a finite mixture model and a Dirichlet prior for the probabilities of the latent classes. A dynamic variant with a first-order Markov model was presented by Yang *et al.* (2011). An extension of this model that relies on infinite mixture models based on the Dirichlet process has been proposed by Kemp *et al.* (2006) and Xu *et al.* (2006). More recently, Airoldi *et al.* (2008) introduced the idea of mixed membership stochastic block models for binary networks wherein the actors can belong to more than one latent class to explore subjects with multiple roles in the network, and Xing *et al.* (2010) developed a temporal extension.

Other approaches to dynamic stochastic block models include the work of Wang *et al.* (2014) who proposed a method for change point detection using hypothesis testing and locality statistics to identify anomalies over time, and the state space model of Xu and Hero (2014) which introduces the extended Kalman algorithm as an alternative to Markov chain Monte Carlo sampling. An extensive review of methods for anomaly detection in dynamic networks was presented by Ranshous *et al.* (2015) including some relevant probabilistic models. For example, Heard *et al.* (2010) utilized Bayesian discrete time counting processes with conditionally independent increments to identify nodes whose relationships have changed over time (see also Robinson and Priebe (2013)), and Perry and Wolfe (2013) considered a multivariate point process to model directed interactions between actors in continuous time, and explored the effect of homophily and network effects in the prediction of future interactions. Recent work presented by Matias and Miele (2017) proposed a frequentist stochastic block model with independent Markov chains for the evolution of the nodes groups, thus yielding a different community structure at each time point. This work overcomes common identifiability issues in other existing modelling approaches by focusing on a prespecified number of groups that are stable over time and show high connectivity within. As pointed out in Matias and Miele (2017), their approach is better suited for contact networks where community stability is an important feature.

In this paper, we propose modelling the dynamics of financial trading networks by using an extension of the Bayesian infinite dimensional model of Kemp *et al.* (2006). The model that we propose is suitable for change point identification by accounting for dependence of the network structure over time through a hidden Markov model (HMM). We also incorporate more general hierarchical priors on the interaction probabilities as well as the partition structure. In particular, we allow for disassortative community behaviour to provide a more flexible representation of trading patterns in the network over time. In addition, the number of communities is not specified in advance but rather estimated from the data at each state of the network. By assuming a non-parametric prior in the partition structure, the model enables the distributions of the number of groups and group sizes to adapt accordingly in the presence of market perturbations. We expect the flexibility of disassortative behaviour and the community distributions over time to aid the change point detection task significantly, which is the main focus in this paper. In finance, regime switching models have been used in many contexts such as applications to model stock returns (Guidolin and Timmermann, 2005; Kim *et al.*, 2001; Perez-Quiroz and Timmermann, 2000), in asset allocation (Ang and Bekaert, 2002a), business cycles (Filardo, 1994) and interest rates (Ang and Bekaert, 2002b) where prediction is also a key objective. As we show in our illustration, by developing a flexible dynamic, fully probabilistic model for array-valued data we can monitor structural changes in market microstructure while making short-term predictions of future trading relationships.

2. Data

The data that we analyse in this paper consist of proprietary transactions made by traders in the New York Mercantile Exchange (NYMEX) natural gas futures market between January 2005 and December 2008. A total of 970 unique traders participated in proprietary transactions at least once over the 4 years to December 2008. However, this list includes traders who either abandoned proprietary trading or went bankrupt during the period under study, as well as traders who entered the market after January 2005. Indeed, only between 240 and 340 traders participated in trades each week (Fig. 1(a)). Since identifying trading strategies in the presence of the noise that is introduced by these transient traders is difficult and we have no detailed information about the times at which different traders entered or left the market outside the 4-year period, our analysis focuses on a total of 290 traders who participated in transactions in at least 25 weeks (out of 201) and had a number of trading partners of at least 50 over the 4-year period. The trading network with the selected 290 traders is a strongly connected graph if we take into account all the transactions over the whole time period. Note that traders were anonymized and are identified in the paper by using numbers.

We used the transaction data to construct weekly trading networks where a link from trader A to trader B was established if there was at least one transaction during that week in which A was the seller and B was the buyer. Data were grouped weekly because this is a low liquidity market in which daily transactions do not provide any strong signal of community behaviour and the number of daily participants is too low compared with the total number of traders who were involved in the market over the 4-year period. Monthly transactions are coarser than weekly transactions but show similar patterns and weekly observations enable us to have a more refined exploration of the network data. In contrast, we considered binary networks instead of weighted networks (e.g. the number or volume of transactions) because the presence or absence of links provides enough information to understand the dynamic of the network in terms of trader partnerships and community patterns.

Fig. 1 presents time series plots of the mean total degree (which measures the total number of links that a trader has), clustering coefficients (which measure the tendency of traders to establish transitive relationships) and assortativity coefficients (which measure the tendency of traders to interact with other traders who are similar to themselves) for the 201 networks in the NYMEX data set. These plots suggest at least a couple of change points in the structure of the network, including one around September 5th, 2006 (which corresponds to the date of introduction of electronic trading in this market via the Chicago Mercantile Exchange Globex platform). To investigate the presence of change points in more detail we fitted a Bayesian HMM with bivariate Gaussian emissions (see Appendix A for details of the model). First we fitted the

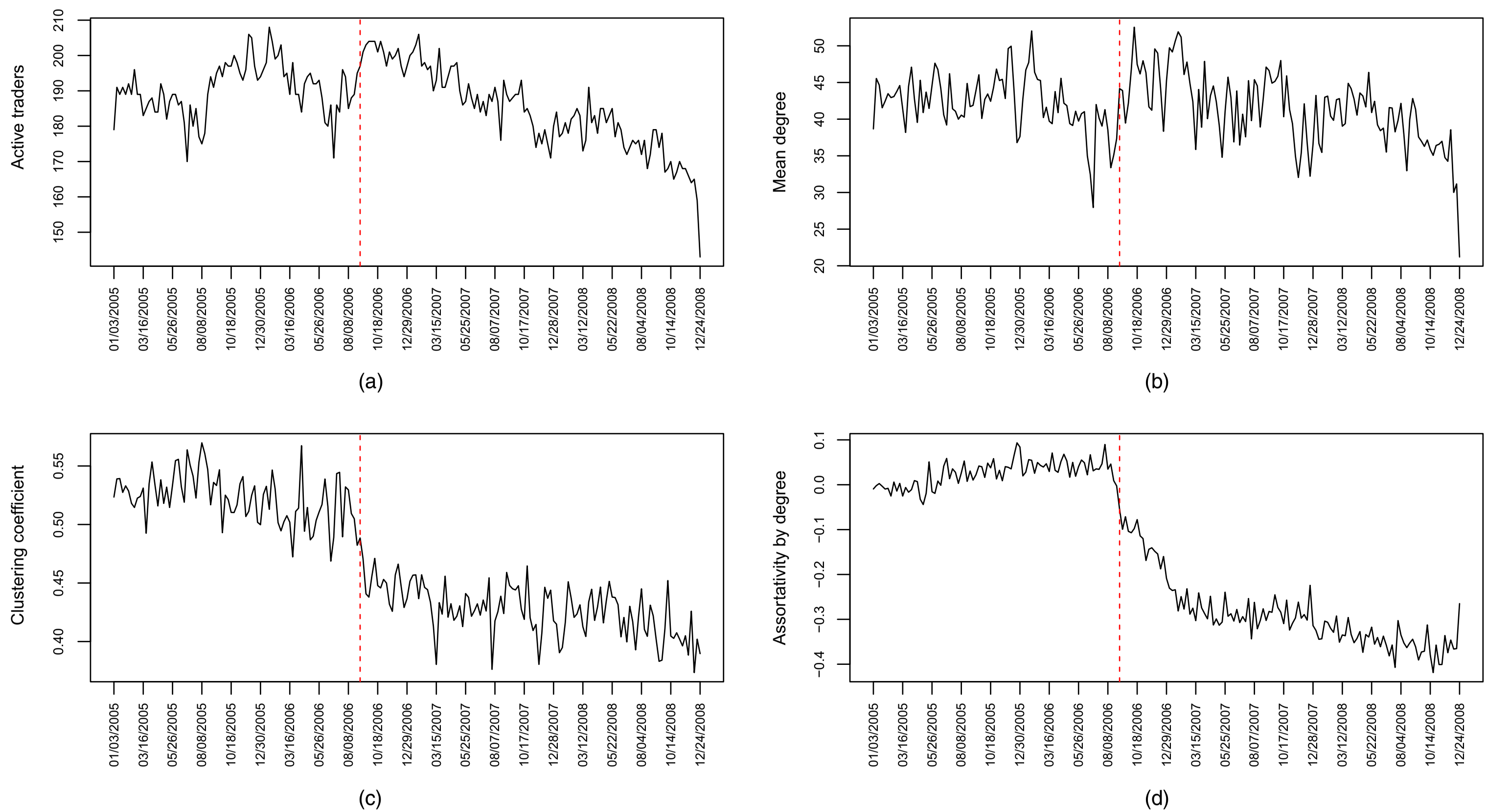
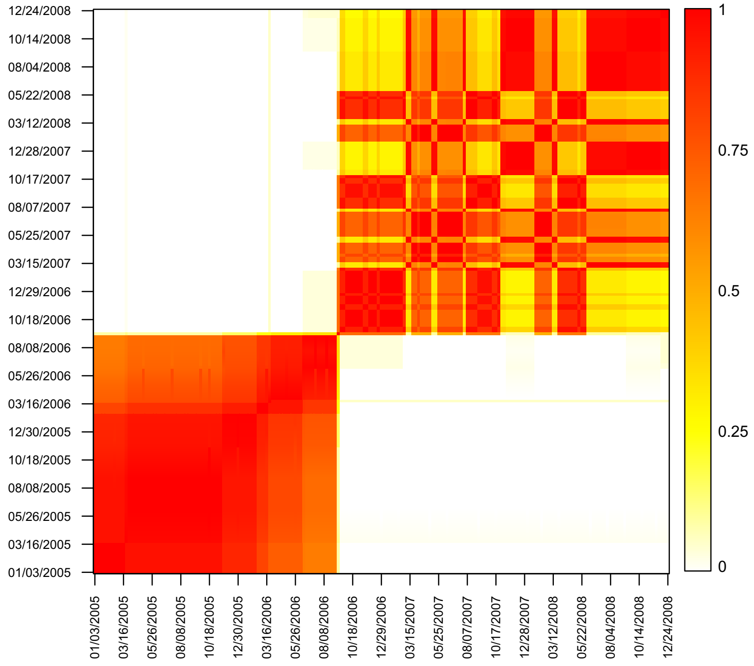
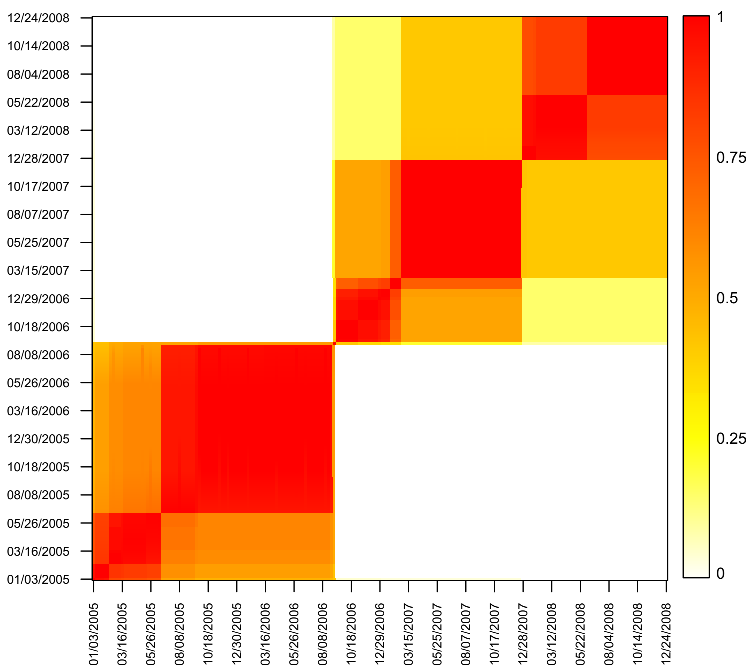


Fig. 1. (a) Number of active traders, (b) mean total degree and (c) clustering and (d) assortativity coefficients for weekly trading networks in the NYMEX natural gas futures market between January 2005 and December 2008: ⁝, September 5th, 2006, when electronic trading was introduced in this market



(a)



(b)

Fig. 2. Mean posterior pairwise incidence matrix for the NYMEX networks under a simple HMM with Gaussian emissions: (a) results based on the mean total degree and the clustering coefficient; (b) results based on the clustering and assortativity coefficients

model to the bivariate time series of mean total degree and clustering coefficient, and then to the bivariate time series of clustering and assortativity coefficients. Fig. 2 shows the marginal posterior probability that any pair of weeks are assigned to the same latent state on each model. These graphs illustrate that the analysis of networks based on summary statistics depends substantially on the *ad hoc* choice of the summaries. Indeed, although both graphs provide evidence of a change point around early September 2006, they disagree on whether other change points are present and, if so, when those happened.

Finally, Fig. 3 presents a matrix representation of the trading network that is associated with the week of February 22nd, 2005 (traders have been reordered to make the graph easier to read). The graph suggests groups of traders who are structurally equivalent, including a large group of inactive and low activity traders who do not participate or have a low number of transactions during this particular week, as well as several small groups of traders including a couple with a high number of intragroup and a relatively low number of intergroup transactions. This suggests that a stochastic block model might be a reasonable model for individual trading networks.

3. Modelling approach

3.1. Stochastic block models for financial trading networks

We encode a financial trading network among n traders by using an $n \times n$ binary sociomatrix $\mathbf{Y} = (y_{i,j})$, where $y_{i,j} = 1$ if trader i sold at least one contract to trader j, and $y_{i,j} = 0$ otherwise. Since we focus on proprietary trading (i.e. transactions carried out by the traders with their own money, rather than their clients'), we adopt the convention $y_{i,i} \equiv 0$, as traders do not

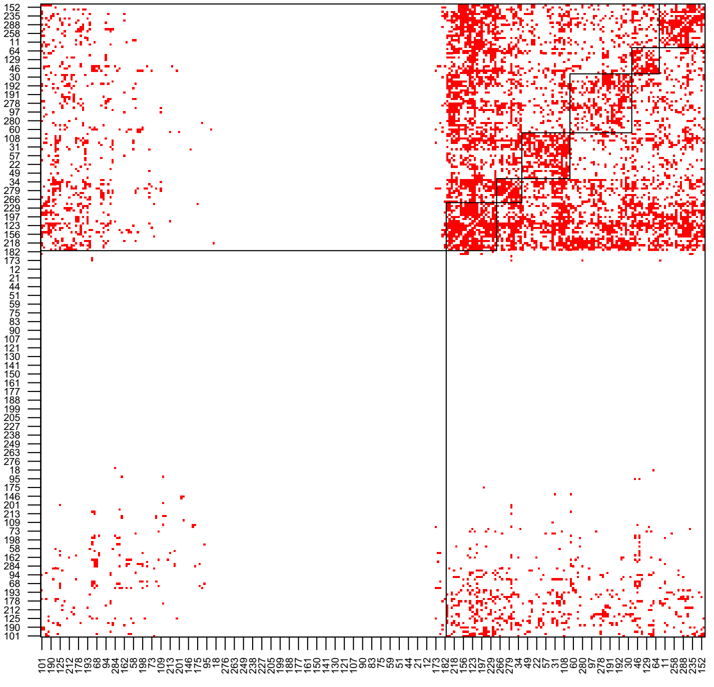


Fig. 3. Sociomatrix for the trading network associated with the week of February 22nd, 2005: ——, |, one possible partition of the traders into groups of structurally equivalent nodes

buy from themselves. Note that treating the network as binary ignores information about the transactions such as the number, maturities and prices of the contracts. We proceed in this way for two reasons. First, in some markets (i.e. black pools) the prices and number of contracts might not be disclosed, making it impossible to apply more general models. Second, even if available, this extra information provides limited additional information about the identity of counterparties subject to contagion risks. Nonetheless, the framework that we describe here can be easily extended to more general types of weighted networks.

A stochastic block model for $\mathbf{Y}$ assumes that its entries are conditionally independent given two sets of parameters: a vector of discrete indicators $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_n)$, where $\xi_i=k$ if and only if trader i belongs to community $k=1,\ldots,K$, and a $K\times K$ matrix $\boldsymbol{\Theta}=(\theta_{k,l})$ such that $\theta_{k,l}$ represents the probability that a member of community k sells a contract to a member of community l. Therefore,

$$y_{i,j}|\boldsymbol{\xi},\boldsymbol{\Theta}\sim \mathrm{Ber}(\theta_{\xi_i,\xi_j}).$$

K represents the maximum potential number of trading communities allowed *a priori*. The effective number of trading communities K^* in the sample could potentially be smaller than K.

A Bayesian formulation for this model is completed by eliciting prior distributions for K, $\boldsymbol{\xi}$ and $\boldsymbol{\Theta}$. In what follows we set $K=\infty$ and let the indicators be independent *a priori* where

$$\Pr(\xi_i=k|\mathbf{w})=w_k, \qquad k=1,2,\ldots,$$

and the vector of weights $\mathbf{w}=(w_1,w_2,\ldots)$ is constructed so that

$$w_k=v_k\prod_{s<k}(1-v_s), \qquad v_k\sim \mathrm{beta}(1-\alpha,\beta+\alpha k), \tag{1}$$

for $0\leqslant\alpha<1$ and $\beta>-\alpha$. By setting $K=\infty$, the model allows for the effective number of components K^* to be as large as the number of traders n, for any n.

The formulation in model (1) is equivalent to the constructive definition of the Poisson–Dirichlet process (Pitman, 1995; Pitman and Yor, 1997), with $\alpha=0$ leading to the Dirichlet process. Hence, the implied prior on the effective number of trading communities K^* and the size of those communities, $m_1,\ldots,m_{K^*}$, is given by

$$\frac{\Gamma(\beta+1)}{(\beta+\alpha K^*)\,\Gamma(\beta+n)}\prod_{k=1}^{K^*}(\beta+\alpha k)\frac{\Gamma(m_k-\alpha)}{\Gamma(1-\alpha)}.$$

Larger values of α or β favour *a priori* a larger effective number of communities K^*. Setting $\alpha=0$ leads to the prior expected number of communities to grow logarithmically with n, whereas for $\alpha>0$ the expected number of components grows as a power of the number of traders.

Consider now specifying a prior on the matrix of interaction probabilities $\boldsymbol{\Theta}$. In this case we let

$$\theta_{k,l}|a_{\mathrm{O}},b_{\mathrm{O}},a_{\mathrm{D}},b_{\mathrm{D}}\sim\begin{cases}\mathrm{beta}(a_{\mathrm{O}},b_{\mathrm{O}}), & k\neq l,\\ \mathrm{beta}(a_{\mathrm{D}},b_{\mathrm{D}}), & k=l,\end{cases}$$

where the subscripts D and O denote diagonal and off diagonal respectively.

This prior is more general than those typically used in stochastic block models, as it allows the distribution of the diagonal and off-diagonal elements of $\boldsymbol{\Theta}$ to have different hyperparameters. This ensures additional flexibility in terms of the implied degree distribution of the network, while still ensuring that both $p(\mathbf{Y})$ and $p(\boldsymbol{\Theta})$ are jointly exchangeable, i.e. that the distributions are invariant to the order in which traders or communities are labelled (Aldous, 1981). In addition, it enables us to define an assortative index for the network as

$$\begin{aligned}\Upsilon &= \log\{E(\theta_{k,k}|a_{\mathrm{D}}, b_{\mathrm{D}})\} - \log\{E(\theta_{k,l}|a_{\mathrm{O}}, b_{\mathrm{O}})\}\\ &= \log\left(\frac{a_{\mathrm{D}}}{a_{\mathrm{D}}+b_{\mathrm{D}}}\right) - \log\left(\frac{a_{\mathrm{O}}}{a_{\mathrm{O}}+b_{\mathrm{O}}}\right),\end{aligned}$$

and a cycle-type transitivity index

$$\chi = \Pr(y_{i,j}=1|y_{j,k}=1, y_{k,i}=1, a_{\mathrm{O}}, b_{\mathrm{O}}, a_{\mathrm{D}}, b_{\mathrm{D}}, \alpha, \beta) = \frac{\chi_N}{\chi_{\mathrm{D}}},$$

where

$$\begin{aligned}\chi_N = &\frac{(1-\alpha)(2-\alpha)}{(\beta+1)(\beta+2)} \frac{(a_{\mathrm{D}}+2)(a_{\mathrm{D}}+1)a_{\mathrm{D}}}{(a_{\mathrm{D}}+b_{\mathrm{D}}+2)(a_{\mathrm{D}}+b_{\mathrm{D}}+1)(a_{\mathrm{D}}+b_{\mathrm{D}})}\\ &+3\frac{(1-\alpha)(\beta+\alpha)}{(\beta+1)(\beta+2)} \frac{a_{\mathrm{D}}}{a_{\mathrm{D}}+b_{\mathrm{D}}}\left(\frac{a_{\mathrm{O}}}{a_{\mathrm{O}}+b_{\mathrm{O}}}\right)^2 + \frac{(\beta+\alpha)(\beta+2\alpha)}{(\beta+1)(\beta+2)}\left(\frac{a_{\mathrm{O}}}{a_{\mathrm{O}}+b_{\mathrm{O}}}\right)^3,\end{aligned}$$

and

$$\begin{aligned}\chi_{\mathrm{D}} = &\frac{(1-\alpha)(2-\alpha)}{(\beta+1)(\beta+2)} \frac{(a_{\mathrm{D}}+1)a_{\mathrm{D}}}{(a_{\mathrm{D}}+b_{\mathrm{D}}+1)(a_{\mathrm{D}}+b_{\mathrm{D}})} + 2\frac{(1-\alpha)(\beta+\alpha)}{(\beta+1)(\beta+2)} \frac{a_{\mathrm{D}}}{a_{\mathrm{D}}+b_{\mathrm{D}}} \frac{a_{\mathrm{O}}}{a_{\mathrm{O}}+b_{\mathrm{O}}}\\ &+ \frac{(\beta+\alpha)(\beta+\alpha+1)}{(\beta+1)(\beta+2)}\left(\frac{a_{\mathrm{O}}}{a_{\mathrm{O}}+b_{\mathrm{O}}}\right)^2.\end{aligned}$$

These two indices are model-based alternatives to assortativity by degree and the clustering coefficients that were discussed in Fig. 1 (Rodriguez and Reyes, 2013).

3.2. Hidden Markov models for time series of financial trading networks

We are interested in extending the hierarchical block model that was described in Section 3.1 to model a time series of financial trading networks $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$. The extension is built with two goals in mind. First, we are interested in identifying events that are associated with structural changes in the network and, therefore, in the microstructure of the market. Second, we aim at making short-term predictions about the structure of the network in future periods. For these reasons, we focus our attention on the use of HMMs for network data. HMMs are widely used in financial (for example, see Ryden *et al.* (1998) and references therein) and biological (e.g. Yau *et al.* (2011) and references therein) applications where there is interest in identifying structural changes in the system under study. Hence, they represent a natural alternative in this context.

More specifically, consider now a sequence $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$ of binary trading networks observed over T consecutive time intervals, where all networks are associated with a common set of n traders. In addition, let $\zeta_1, \ldots, \zeta_T$ be a sequence of unobserved state variables such that $\zeta_t = s$ indicates that the market is in state $s \in \{1, 2, \ldots, S\}$ during period $t = \{1, 2, \ldots, T\}$. Each state has associated with it a vector of community indicators $\boldsymbol{\xi}_s = (\xi_{1,s}, \ldots, \xi_{n,s})$ with $\xi_{i,s} \in \{1, 2, \ldots, K\}$ and a matrix of interaction probabilities $\boldsymbol{\Theta}_s = (\theta_{k,l,s})$ representing respectively the grouping of traders into trading communities and the probabilities of trades between communities when the system is in state s. Analogously to our previous discussion, S and K represent the maximum number of states and the maximum number of trading communities that are allowed by the model *a priori*. The effective number of states S^* and the effective number of communities on each state $K_1^*, \ldots, K_S^*$ are potentially smaller than S and K respectively.

Conditionally on the state parameters, observations are assumed to be independent, i.e.

$$y_{i,j,t}|\zeta_t, \{\boldsymbol{\xi}_s\}, \{\boldsymbol{\Theta}_s\} \sim \mathrm{Ber}(y_{i,j,t}|\theta_{\xi_{i,\zeta_t}, \xi_{j,\zeta_t}, \zeta_t}).$$

Hence, the joint likelihood for the data can be written as

$$p(\{\mathbf{Y}_t\}|\{\zeta_t\},\{\boldsymbol{\xi}_s\},\{\boldsymbol{\Theta}_s\}) = \prod_t^T \prod_{i=1}^n \prod_{\substack{j=1\\ j\neq i}}^n \theta_{\xi_{i,\zeta_t},\xi_{j,\zeta_t},\zeta_t}^{y_{i,j,t}} (1-\theta_{\xi_{i,\zeta_t},\xi_{j,\zeta_t},\zeta_t})^{1-y_{i,j,t}}$$
$$= \prod_{s=1}^S \prod_{k=1}^K \prod_{l=1}^K \prod_{(i,j,t)\in A_{k,l,s}} \theta_{k,l,s}^{y_{i,j,t}} (1-\theta_{k,l,s})^{1-y_{i,j,t}},$$

where $A_{k,l,s} = \{(i,j,t): i\neq j, \zeta_t = s, \xi_{i,\zeta_t} = k, \xi_{j,\zeta_t} = l\}$ is the set of observations that are associated with the interactions between communities k and l in state s.

To account for the persistence in network structure, we assume that the evolution of the system indicators follows a first-order Markov process with transition probabilities

$$p(\zeta_t = s|\zeta_{t-1} = r, \{\boldsymbol{\pi}_r\}) = \pi_{r,s},$$

where $\boldsymbol{\pi}_r = (\pi_{r,1},\ldots,\pi_{r,S})$, the rth row of the transition matrix $\boldsymbol{\Pi} = (\pi_{r,s})$, must satisfy $\Sigma_{s=1}^S \pi_{r,s} = 1$. A natural prior for $\boldsymbol{\pi}_r$ is a symmetric Dirichlet distribution:

$$\boldsymbol{\pi}_r|\gamma \sim \mathrm{Dir}\left(\frac{\gamma}{S},\frac{\gamma}{S},\ldots,\frac{\gamma}{S}\right).$$

As $S\to\infty$, the induced distribution of transitions over states is equivalent to that generated by a Dirichlet process prior with concentration parameter γ (for example, see Green and Richardson (2001)). Therefore the model is similar in spirit to the infinite HMM that was discussed in Rodriguez (2012) (see also Teh *et al.* (2006)). However, our construction does not couple the values of $\boldsymbol{\pi}_1,\boldsymbol{\pi}_2,\ldots$ through a common centring probability. This is in contrast with the infinite HMM, where all transition probabilities into the same state are assumed to be similar. Indeed, the structure of the infinite HMM model implies that, if state s is highly persistent (i.e. $\pi_{s,s}$ is close to 1), then the probability of transitioning from any other states into state s will also tend to be large: a property that is unappealing when modelling financial trading networks. Since γ plays an important role in controlling the number of effective states S^*, its value is estimated from the data by assigning an exponential prior to it and carrying out a sensitivity analysis to evaluate the effect of our prior choice on model performance.

The specification of the model is completed by eliciting hierarchical priors on the state-specific parameters $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_S$ and $\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_S$. Following the discussion in Section 3.1, we let

$$\Pr(\xi_{i,s} = k|\mathbf{w}_s) = w_{k,s}, \qquad k = 1,2,\ldots,$$

where $w_{k,s} = v_{k,s}\Pi_{h<k}(1-v_{h,s})$ are weights that are constructed from a sequence $v_{1,s}, v_{2,s},\ldots$ where $v_{k,s}\sim \mathrm{beta}(1-\alpha_s, \beta_s + k\alpha_s)$. Again, since the hyperparameters α_s and β_s play a critical role in controlling the number of expected trading communities, they are assigned independent hyperpriors $\alpha_s \sim p(\alpha_s)$ and $\beta_s \sim p(\beta_s)$. A natural choice is to assign α_s a uniform prior on the unit interval and β_s an exponential prior, while carrying out a sensitivity analysis that involves priors that favour small values of α_s as well as priors that favour both lower and higher values for β_s.

Similarly, the interaction probabilities are assigned priors

$$\theta_{k,l,s}|a_{s,\mathrm{O}}, b_{s,\mathrm{O}}, a_{s,\mathrm{D}}, b_{s,\mathrm{D}} \sim \begin{cases} \mathrm{beta}(a_{s,\mathrm{O}}, b_{s,\mathrm{O}}), & k\neq l, \\ \mathrm{beta}(a_{s,\mathrm{D}}, b_{s,\mathrm{D}}), & k = l, \end{cases}$$

where $\{a_{s,\mathrm{O}}\}$, $\{b_{s,\mathrm{O}}\}$, $\{a_{s,\mathrm{D}}\}$ and $\{b_{s,\mathrm{D}}\}$ are independent and gamma distributed with shape parameter c and unknown rates d_{O}, e_{O}, d_{D} and e_{D}, which are in turn assigned exponential priors with means λ_d and λ_e.

4. Computation

The posterior distribution that is associated with our HMM for stochastic block models is not analytically tractable. Therefore, we implemented a Markov chain Monte Carlo algorithm (Robert and Casella, 2005) that simulates a dependent sequence of random draws from the target distribution. Given initial values for the parameters, these are successively updated from their full conditional distributions. Standard Markov chain theory ensures that, after an appropriate burn-in, the values of the parameters that are generated by the algorithm are approximately distributed according to the posterior distribution. To derive the algorithm, we rely on the fact that the joint posterior distribution can be factorized as

$$p(\{\boldsymbol{\Theta}_s\}|\{\boldsymbol{\xi}_s\},\{\zeta_t\},\{a_{s,\mathrm{O}}\},\{b_{s,\mathrm{O}}\},\{a_{s,\mathrm{D}}\},\{b_{s,\mathrm{D}}\},\{\mathbf{Y}_t\}) \\ \times p(\{\boldsymbol{\xi}_s\},\{\zeta_t\},\{a_{s,\mathrm{O}}\},\{b_{s,\mathrm{O}}\},\{a_{s,\mathrm{D}}\},\{b_{s,\mathrm{D}}\},d_{\mathrm{O}},e_{\mathrm{O}},d_{\mathrm{D}},e_{\mathrm{D}},\{\alpha_s\},\{\beta_s\},\gamma|\{\mathbf{Y}_t\}). \tag{2}$$

Since the values of $\theta_{k,l,s}$ are conditionally independent *a posteriori* given the observations, the indicators $\{\zeta_t\}$ and $\{\xi_{i,s}\}$, and the prior parameters $\{a_{s,\mathrm{O}}\}$, $\{b_{s,\mathrm{O}}\}$, $\{a_{s,\mathrm{D}}\}$ and $\{b_{s,\mathrm{D}}\}$, the first term in expression (2) is easy to sample from. Furthermore, conditionally on the other parameters in the model, the state indicators $\zeta_1,\ldots,\zeta_T$ are sampled jointly by using a forward–backward algorithm (Rabiner and Juang, 1986), whereas the full conditional distribution for each collection of indicators $\xi_{1,s},\ldots,\xi_{n,s}$ is sampled by using a collapsed (marginal) Gibbs sampler (Neal, 2000). Details of the algorithm are discussed in Appendix B.

Given a sample from the previous Markov chain Monte Carlo algorithm,

$$(\{\boldsymbol{\Theta}_s^{(b)}\},\{\boldsymbol{\xi}_s^{(b)}\},\{\zeta_t^{(b)}\},\{a_{s,\mathrm{O}}^{(b)}\},\{b_{s,\mathrm{O}}^{(b)}\},\{a_{s,\mathrm{D}}^{(b)}\},\{b_{s,\mathrm{D}}^{(b)}\},d_{\mathrm{O}}^{(b)},e_{\mathrm{O}}^{(b)},d_{\mathrm{D}}^{(b)},e_{\mathrm{D}}^{(b)},\{\alpha_s^{(b)}\},\{\beta_s^{(b)}\},\gamma^{(b)}), \\ b=1,\ldots,B,$$

obtained after an appropriate burn-in period, point and interval estimates for model parameters can be easily obtained by computing the empirical mean and/or the empirical quantiles of the posterior distribution. Note that the model has label switching issues that are associated with the state and community indicators. This produces identifiability problems that can be addressed by focusing inference on identifiable functions of the parameters. For example, posterior coclustering probabilities, $\omega_{t,t'}=\Pr(\zeta_t=\zeta_{t'}|\{\mathbf{Y}_t\})$ are estimated as

$$\hat{\omega}_{t,t'}=\Pr(\zeta_t=\zeta_{t'}|\{\mathbf{Y}_t\})\approx\frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(\zeta_t^{(b)}=\zeta_{t'}^{(b)}),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. The estimates are arranged into a coclustering matrix $(\hat{\omega}_{t,t'})$, which is in turn used to identify the state of the system at each time period through the decision theoretic approach that is used to perform Bayesian model-based clustering presented by Lau and Green (2007). More specifically, the point estimate for the grouping of networks in state s is found by minimizing the posterior expected loss that is associated with a loss function that reflects the penalty that is incurred by pairwise misclassification. The point estimate of the states is found on the basis of the posterior coclustering probabilities $\hat{\omega}_{t,t'}$ and a constant value $0\leqslant K_0\leqslant 1$ usually chosen to be greater than 0.5 to obtain a refined partition with low uncertainty within each cluster. A similar procedure is carried out to obtain the community structure, using the vector of community indicators $\boldsymbol{\xi}_s=(\xi_{1,s},\ldots,\xi_{n,s})$ to estimate the pairwise posterior probability of any two traders to belong to the same community at each state of the system.

The samples from the posterior distribution can also be used as the basis for prediction. For this, note that the probability that trader i sells at least one security to trader j in the unobserved period $T+1$ can be estimated by

$$E(y_{i,j,T+1}|\{\mathbf{Y}_t\}) \approx \frac{1}{B}\sum_{b=1}^{B} \pi^{(b)}_{\zeta_T^{(b)},s}\,\theta^{(b)}_{\xi^{(b)}_{i,s},\xi^{(b)}_{j,s}}.$$

Using a simple 0–1 utility function, a future sell trade from trader i to trader j is predicted as $\hat{y}_{i,j,T+1} = \mathbb{I}[E(y_{i,j,T+1}|\{\mathbf{Y}_t\}) > f]$, for some threshold f that reflects the relative cost that is associated with false positive and false negative links.

5. Analysis for the New York Mercantile Exchange natural gas futures market

In this section we analyse the sequence of $T=201$ weekly financial trading networks from transactions between 290 traders in the NYMEX natural gas futures market that was introduced in Section 2. The results that are presented in this section are based on 50000 iterations after a burn-in period of 50000 iterations with a thinner of 5. Convergence of the algorithm was diagnosed by using the R-statistic of Gelman and Rubin (1992) for two independent chains of the log-likelihood function and by a visual evaluation of trace plots (Appendix C). We also monitored the number of active states S^* and the mean and variance over time of the assortativity and transitivity indices $\{\Upsilon_t\}$ and $\{\chi_t\}$. In terms of hyperparameters, the maximum number of states is set to $S=30$, the prior means for γ and $\{\beta_s\}$ are assigned exponential priors with unit mean and the priors for d_{O}, e_{O}, d_{D} and e_{D} are exponential distributions with mean 2. This specification implies that, *a priori*, $E(\Upsilon_t)=0$ for all $t=1,\ldots,T$, so we favour neither assortative nor disassortive trading communities *a priori*.

5.1. Identifying changes in market microstructure

The posterior estimate of the coclustering matrix for the latent states $\zeta_1,\ldots,\zeta_T$ is presented in Fig. 4, along with a point estimator for the grouping of networks into states (recall Section 4). This point estimator (obtained for a value $K_0=0.7$) suggests that the structure of the trading networks alternates between five states. The first state runs between early January 2005 and late February 2006. The second state runs between late February 2006 and early August 2006, which was about a month before the electronic market is introduced. The third state then lasts until early February 2007 before the system transitions to a new state for a short period of 6 weeks. After that, the system seems to transition to a fifth state in late March 2007 and remains in this state for the remaining weeks in the period observed. Our proposed model is sufficiently flexible to allow for early states to be revisited but in this particular application a simple change point model could have been a viable option. Also, it is clear from the heat map that, although some uncertainty is associated with this point estimate of the system states (mostly in time of the transitions between states), this uncertainty is relatively low. Note that these results have some similarities with those we reported in Fig. 2, but also some important differences. In particular, all models agree on the presence of a change point associated with the introduction of electronic trading on September 5th, 2006, but disagree on the timing and structure of other change points.

Fig. 5 shows estimates of the community structure that is associated with representative weeks from each of the five states. In this case, the values of the incidence matrices represent the

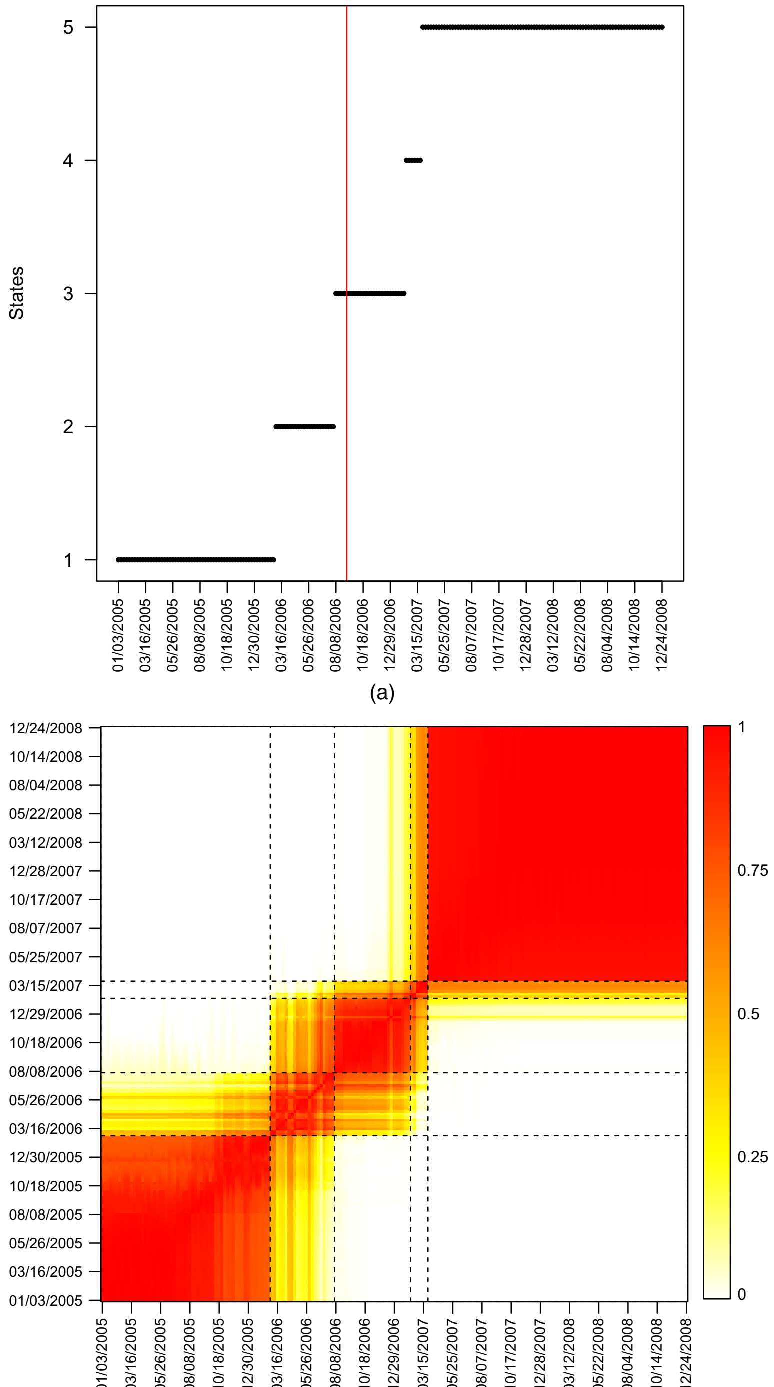


Fig. 4. (a) Point estimate of the states for the 201 weeks observed for the trading network (|, introduction of the electronic platform on week 85) and (b) mean posterior pairwise incidence matrix for the NYMEX networks under our block model HMM, illustrating the uncertainty associated with this point estimate

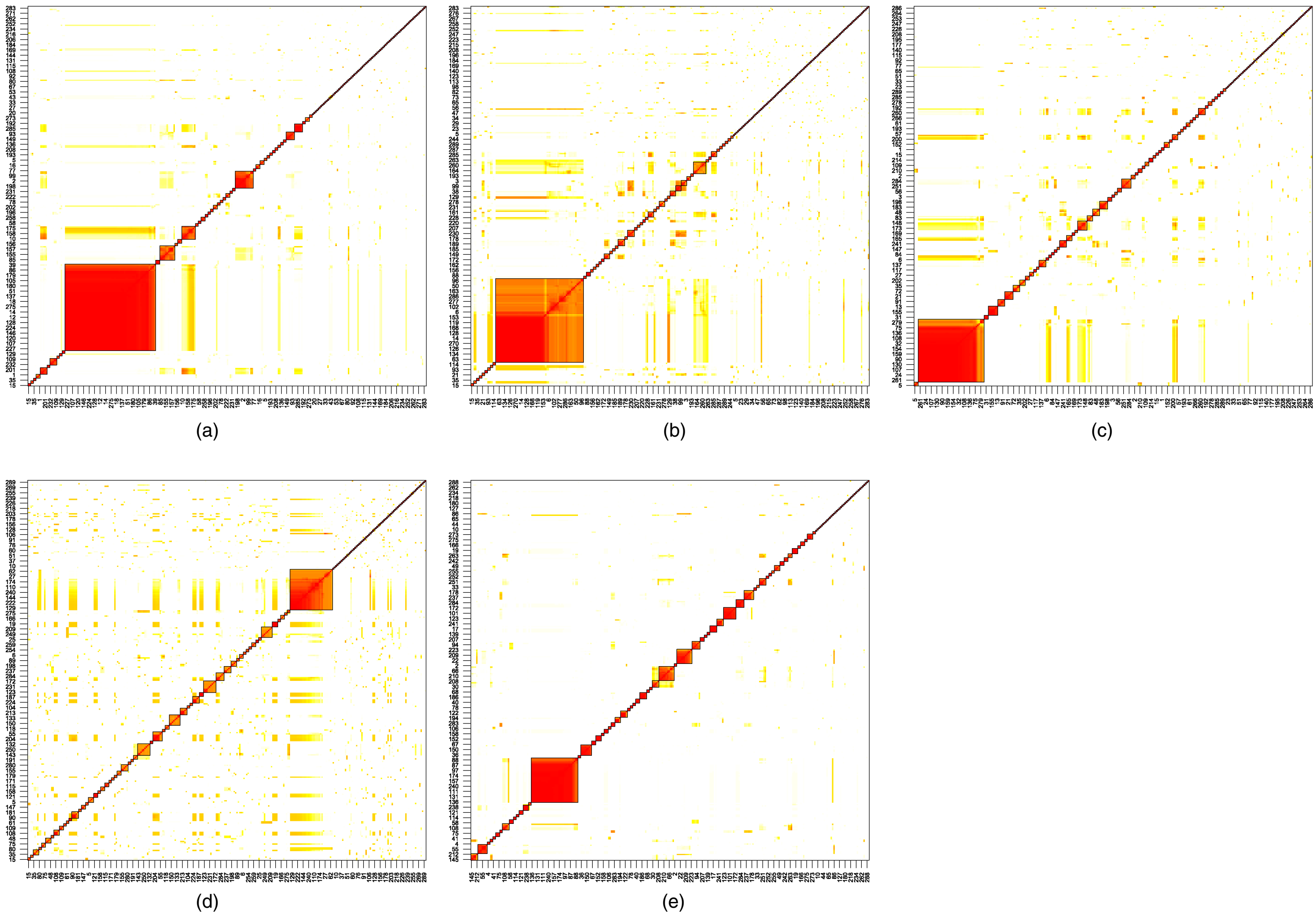


Fig. 5. Mean posterior pairwise incidence matrices of traders for weeks (a) 40, (b) 65, (c) 90, (d) 110 and (e) 120 representative from states 1–5 respectively: the matrices are arranged according to the optimal clustering for each state of the system

posterior probability of any two traders to belong to the same cluster under each state of the system. The matrices are arranged according to the optimal clustering for each state to showcase the community structure over time. Although there are some similarities, the overall structure of the communities is quite different. For example, state 1 is characterized by a large group of 66 mostly inactive traders, seven small communities comprising between five and 13 traders each, whereas all other traders tend to fall, for the most part, into singleton clusters or clusters of size 2. In contrast, although state 4 also exhibits some singleton clusters, it also shows a larger number of small communities comprising between four and 11 traders each.

From Fig. 5 we see low uncertainty about the membership of the traders to the communities and a big number of singletons for all states. Because of the amount of information that is available for the traders over the 201 weeks, the model produces a very refined profile for the role of the traders in the communities leading to a large number of groups with only one element.

The time series plots for the estimates of the assortativity and transitivity indices $\Upsilon_1, \ldots, \Upsilon_T$ and $\chi_1, \ldots, \chi_T$ are presented in Fig. 6. Recall that these quantities are model-based alternatives to the assortativity by degree and the clustering coefficient that is presented in Fig. 1. Both sets of plots share some common features, revealing mild assortativity and higher transitivity before September 2006 and disassortative networks with lower transitivity afterwards. This makes sense because we would expect that the introduction of an electronic market would limit the effect of social connections between traders (which tend to be assortative and transitive) and favour connections based on differential trending strategies (which tend to be disassortative).

5.2. Network prediction

As we discussed in Section 1, besides identifying change points in market microstructure, a secondary goal is to predict future trading partnerships. To assess the predictive abilities of the model we ran an out-of-sample cross-validation exercise where we held out the last 10 weeks in the data set and made one-step-ahead predictions for the structure of the held-out networks. More specifically, for each $t = 191, 192, \ldots, 200$ we use the information that is contained in $\mathbf{Y}_1, \ldots, \mathbf{Y}_t$ to estimate the model parameters and obtain predictions for $\hat{\mathbf{Y}}_{t+1}$ for different values of the threshold f. Each of these predictions is compared against the observed network $\mathbf{Y}_{t+1}$, the number of false and true positive results is computed and a receiver operating characteristic curve is constructed. For comparison, the same exercise was performed with a temporal ERGM and the dynamic stochastic block model dynsbm that was proposed in Matias and Miele (2017). We used the `xergm` and `dynsbm` packages in R to estimate the temporal ERGM (Leifeld *et al.*, 2014) and the the dynamic stochastic block model respectively. More specifically, the temporal ERGM is estimated with the `btergm` function, which implements the bootstrapped pseudolikelihood procedure that was presented by Desmarais and Cranmer (2012). The model that we fit includes all the typical ERGM terms, the square root of in- and out-degrees as node covariates and the lagged network and the delayed reciprocity to model cross-temporal dependences.

In contrast with our proposed HMM, the implementation of the model of Matias and Miele (2017) requires the number of groups to be specified in advance. We computed the integrated classification likelihood for values between 3 and 20 of the number of groups by using the `dynsbm` package (see Fig. 10 in Appendix D). According to these results the optimal number of groups is 8. Fig. 7 displays the group membership of the 290 traders over the 4-year period. The traders were arranged according to their group membership in the first week to show the stability of the community structure over time. From these results we observe that the community membership varies considerably each week, thus making the identification of structural changes in the evolution of the network over time challenging. Before the introduction of the

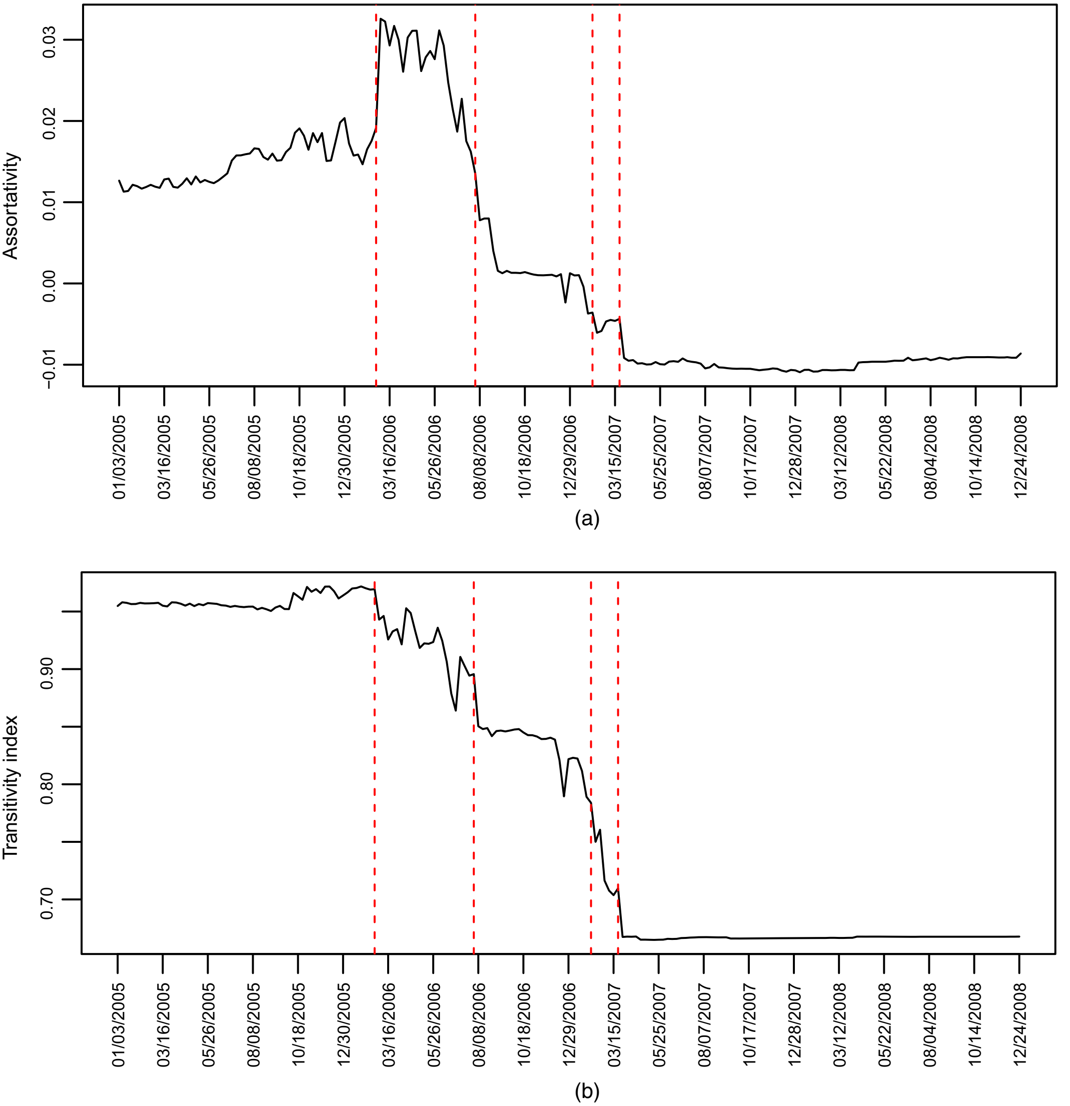
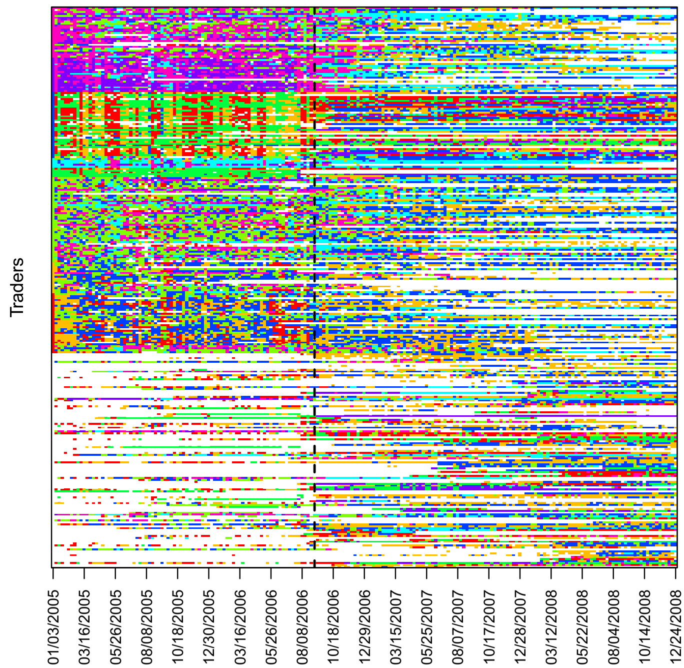
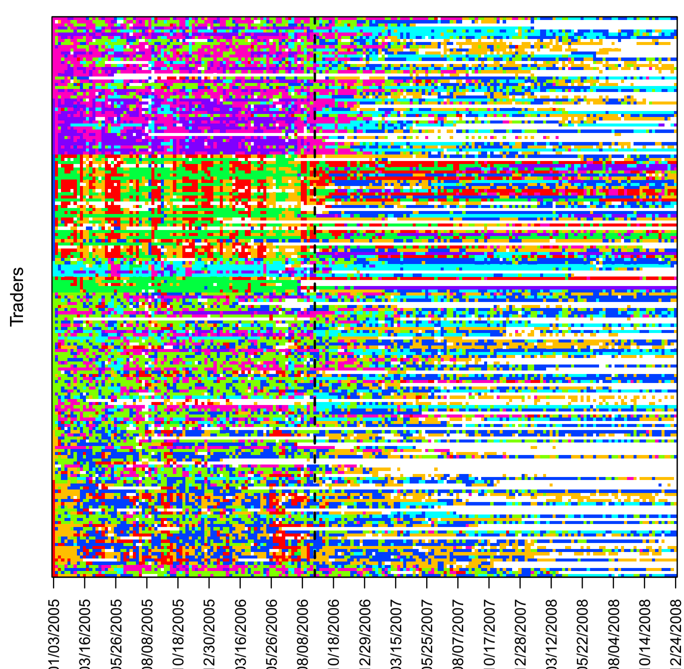


Fig. 6. Time series plot for (a) assortativity and (b) transitivity indices: ⁝, transitions across states identified from Fig. 4

electronic platform, we see how the community structure in the market was relatively more stable. Furthermore, a fair amount of traders who were consistently active before the introduction of electronic trading became less active or exited the market completely and vice versa (see Fig. 7(b)). These results are consistent with our initial exploratory analysis of the data and the inference performed with our model. Fig. 8 displays the ROC curves that are associated with one-step-ahead out-of-sample predictions from our HMM and a comparison of area under the curve, AUC, values for the predictions of the last 10 weeks for the three models considered. The results indicate that the prediction ability of all the models is very good with an average AUC of 95%, 94% and 92% for dynsbm, our proposed HMM and the temporal ERGM respectively. Even though in this particular case dynsbm slightly outperforms our model (with the exception



(a)



(b)

Fig. 7. (a) Group membership for the 290 traders in the NYMEX market over the 4-year period obtained from dynsbm (the rainbow colours represent the eight communities and the colour white indicates that the trader was inactive that particular week) and (b) a close-up of (a) excluding the inactive traders (white) in the first week (ı, introduction of electronic trading in the market)

of weeks 198 and 200) the results are encouraging considering that our HMM model is more parsimonious (only five community configurations or states), and the community structure is more interpretable and suitable for change point detection.

5.3. Sensitivity analysis

To assess the effect of our prior choice on posterior inference we conducted a sensitivity analysis where the model was fitted with somewhat different priors. In particular, we used independent beta priors with mean $\frac{1}{10}$ and variance 9/1100 for each α_s, as well as exponential priors with means $\frac{1}{3}$ and 3 for each β_s. In contrast, exponential priors with mean 2 were also used for d_{O}, e_{O}, d_{D} and e_{D}. Although inferences on the community structure were somewhat affected by prior choices and inferences on the state parameters as well as the assortativity and transitivity indices and the predictive performance were essentially unchanged.

6. Discussion

We have presented a class of HMMs for financial trading networks that have clear potential for market regulatory oversight. Key applications of these models include identifying specific events (such as large trader failures or specific changes in market rules) that affect market stability, as well as identifying frequent trading counterparties that might be likely collusion partners or particularly at risk in case of bankruptcies. The financial trading network in the NYMEX market displays patterns of disassortative behaviour and drastic structural changes after the introduction of electronic trading, which also affected traders' survival.

Although the use of an HMM enables us to account for time dependence and is useful for identifying structural changes in the system, a structure that assumes abrupt changes in the network might be too restrictive for predictive purposes. Models based on fragmentations and coagulations (for example, see Bertoin (2006)) that allow for smooth evolution in the community structure can potentially enable improved predictions. Recent work on dynamic network models that consider parameters of link (and non-link) persistence over time could also be explored in this context (Xu, 2015; Friel *et al.*, 2016; Barucca *et al.*, 2017). In this paper we have focused on models for binary networks where only the presence or absence of transactions over a week is recorded. However, when other information is available (e.g. the volume of transaction), the model can be easily extended to incorporate it.

Appendix A: Hidden Markov model with bivariate normal emissions

In Section 2 we fitted an HMM with bivariate Gaussian emissions for different pairs of summary statistics on the NYMEX network. In this appendix we provide a detailed formulation of the model.

Let $\mathbf{x}_t = (x_{1,t}, x_{2,t})'$, where $x_{1,t}$ and $x_{2,t}$ are two summary statistics (such as the clustering and assortativity coefficients) of the network observed on week t. We assume that

$$\mathbf{x}_t | \zeta_t^*, \{\boldsymbol{\mu}_s\}, \{\boldsymbol{\Omega}_s\} \sim N(\boldsymbol{\mu}_{\zeta_t^*}, \boldsymbol{\Omega}_{\zeta_t^*}), \qquad t = 1, \ldots, T,$$

where $\boldsymbol{\mu}_s \sim N(\mathbf{d}, \mathbf{D})$ and $\boldsymbol{\Omega}_s \sim \mathrm{IW}(a, \mathbf{B})$ independently for each $s = 1, \ldots, R$ and, as in our other model in this paper, the state indicators satisfy

$$p(\zeta_t^* = s | \zeta_{t-1}^* = r, \{\boldsymbol{\pi}_r^*\}) = \pi_{r,s}^*, \qquad \boldsymbol{\pi}_r | \gamma^* \sim \mathrm{Dir}\left(\frac{\gamma^*}{R}, \ldots, \frac{\gamma^*}{R}\right).$$

For the analysis that was shown in Section 2 we set $\gamma^* = 1$, and set $\mathbf{d}$ to the mean and $\mathbf{D}$ and $\mathbf{B}$ both to the variance–covariance matrix of the observations. The estimates of the pairwise probabilities $\Pr(\zeta_t^* = \zeta_{t'}^* | \{\mathbf{x}_t\})$ were obtained from 10000 iterations (obtained after a burn-in period of 1000 samples) of a

Markov chain Monte Carlo algorithm that alternates through sampling $\{\boldsymbol{\mu}_s\}$, $\{\boldsymbol{\Omega}_s\}$ and $\{\boldsymbol{\zeta}_t^*\}$ from their corresponding full conditional posterior distributions. The details of the algorithm are very similar to that discussed in Appendix B for the HMM with block model emissions.

Appendix B: Computational algorithm for the hidden Markov model with block model emissions

Here, we provide the details of the Markov chain Monte Carlo algorithm that was discussed in Section 4. The algorithm proceeds by updating the model parameters from the following full conditional distributions.

(a) For each $i=1,\ldots,n$ and occupied states s, $\xi_{i,s}=k$ with probability

$$\Pr(\xi_{i,s}=k|\ldots,\mathbf{Y})$$
$$=\begin{cases} (m_k^{-i}-\alpha_s)\displaystyle\prod_{l=1}^{K^*_{s,-i}} \frac{p[\{y_{i,j,t}:(i,j,t)\in A^i_{k,l,s}\}]}{p[\{y_{i,j,t}:(i,j,t)\in A^{-i}_{k,l,s}\}]}\frac{p[\{y_{j,i,t}:(i,j,t)\in A^i_{k,l,s}\}]}{p[\{y_{j,i,t}:(i,j,t)\in A^{-i}_{k,l,s}\}]}, & k\leqslant K^*_{s,-i},\\ (\beta_s+\alpha_s K^*_{s,-i})\displaystyle\prod_{l=1}^{K^*_{s,-i}} p[\{y_{i,j,t}:(j,t)\in A^{-i}_{l,s}\}]p[\{y_{j,i,t}:(j,t)\in A^{-i}_{l,s}\}], & k=K^*_{s,-i}+1,\end{cases}$$

where $K^*_{s,-i}=\max_{j\neq i}\{\xi_{j,s}\}$, $m_k^{-i}=\Sigma_{j\neq i}\mathbb{I}_{(\xi_{j,s}=k)}$,

$$A^{-i}_{k,l,s}=\{(i',j',t):i'\neq j'\neq i,\zeta_t=s,\xi_{i',\zeta_t}=k,\xi_{j',\zeta_t}=l\},$$
$$A^{i}_{k,l,s}=\{(i',j',t):i'=i,\zeta_t=s,\xi_{j',\zeta_t}=l\}\cup A^{-i}_{k,l,s},$$
$$A^{-i}_{l,s}=\{(j,t):j\neq i,\zeta_t=s,\xi_{j,\zeta_t}=l\},$$

and the marginal predictive distribution, $p[\{y_{i,j,t}:(i,j,t)\in A\}]$ is given by

$$\frac{\Gamma(\Sigma_A y_{i,j,t}+a_s)\,\Gamma(|A|+b_s-\Sigma_A y_{i,j,t})}{\Gamma(a_s+b_s+|A|)}\frac{\Gamma(a_s+b_s)}{\Gamma(a_s)\,\Gamma(b_s)},$$

and $|A|$ is the number of elements in A.

(b) Since the prior for $\theta_{k,l,s}$ is conditionally conjugate, we update these parameters for $k,l\in\{1,\ldots,K^*_s\}$ by sampling from

$$\theta_{k,l,s}|\ldots,\mathbf{Y}\sim\mathrm{beta}\left(\sum_{A_{k,l,s}} y_{i,j,t}+a_s, m_{k,l,s}+b_s-\sum_{A_{k,l,s}} y_{i,j,t}\right)$$

for $A_{k,l,s}=\{(i,j,t):i\neq j,\zeta_t=s,\xi_{i,\zeta_t}=k,\xi_{j,\zeta_t}=l\}$ and $m_{k,l,s}=|A_{k,l,s}|$.

(c) Since the prior for the transition probabilities is conditionally conjugate, the posterior full conditional for $\boldsymbol{\pi}_r$, $r=1,\ldots,S,$ is the Dirichlet distribution

$$p(\boldsymbol{\pi}_r|\ldots,\mathbf{Y})=\prod_{s=1}^{S}\pi_{r,s}^{\gamma/S+n_{rs}-1}$$

for $n_{rs}=|\{t:\zeta_{t-1}=r,\zeta_t=s\}|$.

(d) The posterior full conditional of γ is

$$p(\gamma|\ldots,\mathbf{Y})\propto p(\gamma)\prod_{s=1}^{S}\frac{\Gamma(\gamma)}{\Gamma(\gamma+n_s)}\gamma^{L_s}$$

where $n_s=|\{t:\zeta_t=s\}|$ and $L_s=\Sigma_r\mathbb{I}_{n_{s,r}>0}$ for $n_{s,r}=|\{t:\zeta_{t-1}=s,\zeta_t=r\}|$. Since this distribution has no standard form, we update γ by using a random-walk Metropolis–Hastings algorithm with symmetric log-normal proposal:

$$\log(\gamma^{(p)})|\gamma^{(c)}\sim N\{\log(\gamma^{(c)}),\kappa_\gamma^2\}$$

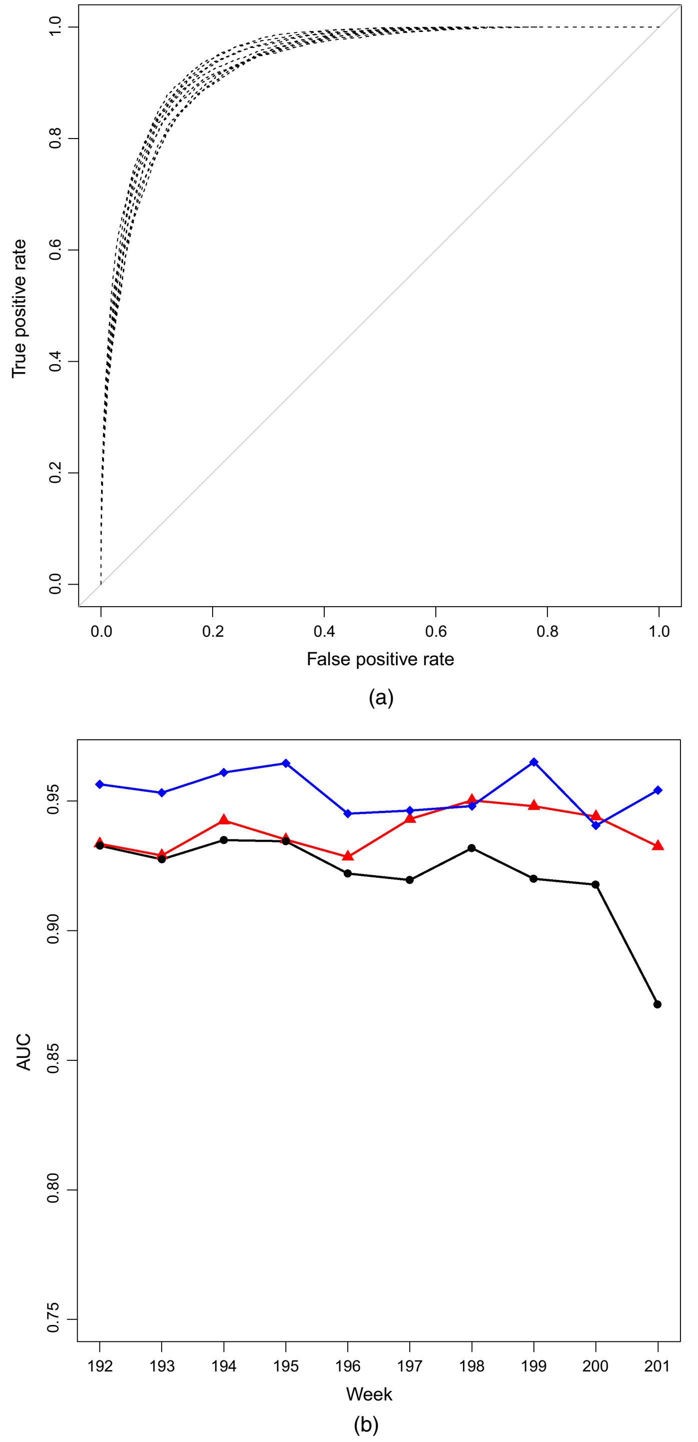


Fig. 8. (a) 10 operating characteristic curves associated with one-step-ahead out-of-sample predictions from our HMM and (b) time series plot of the area under the curves, AUC, for the temporal ERGM (●), dynsbm (◆) and our proposed HMM (▲)

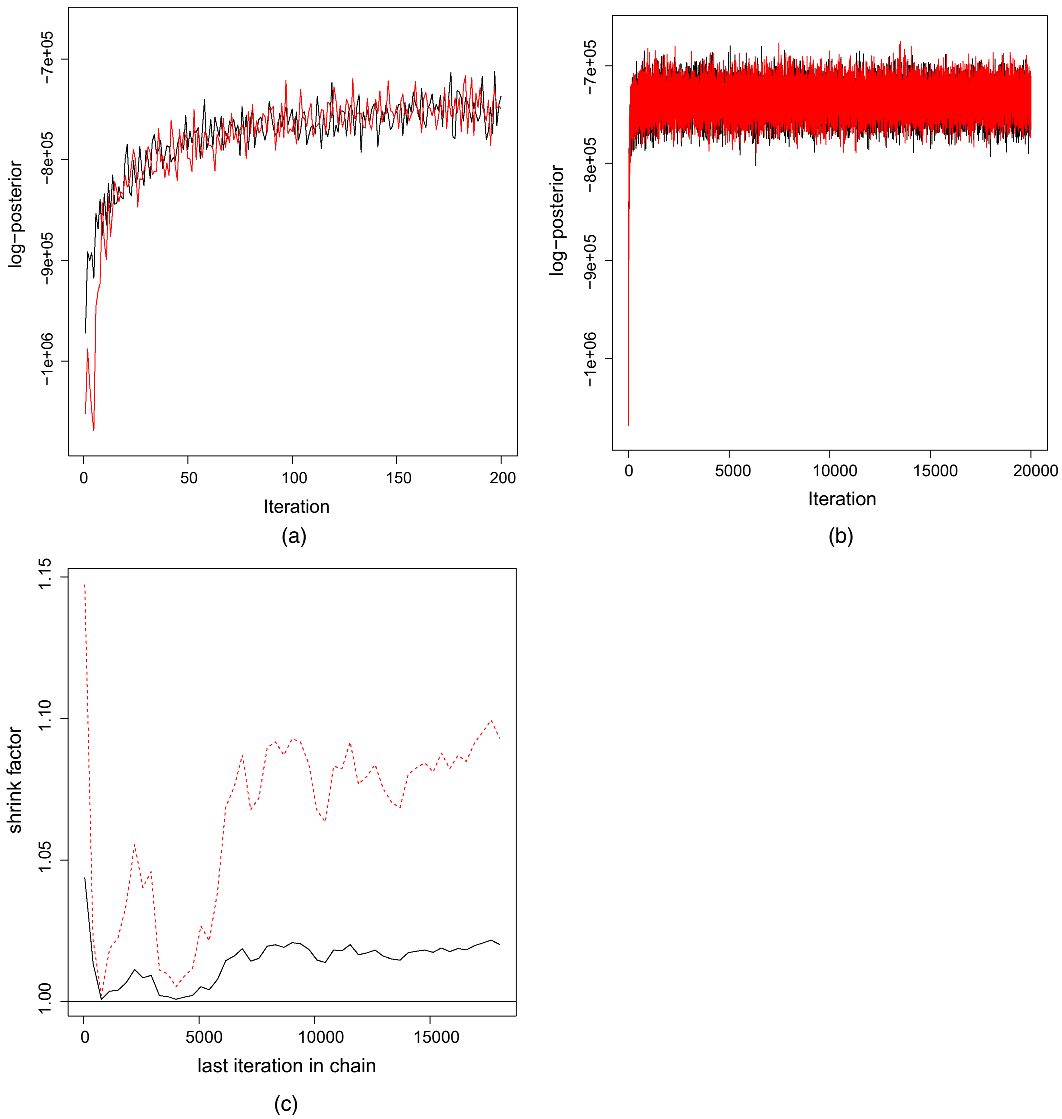


Fig. 9. (a), (b) Trace plots for two Markov chain Monte Carlo chains of the log-posterior with different initial values and (b) Gelman and Rubin's *R*-statistic: ———, median; - - - - -, 97.5% percentile

where κ_γ^2 is a tuning parameter chosen to obtain an average acceptance rate between 30% and 40%.

(e) The posterior full conditional of the pairs $(a_{s,\mathrm{O}}, b_{s,\mathrm{O}})$ and $(a_{s,\mathrm{D}}, b_{s,\mathrm{D}})$ has the following general form:

$$p(a_s, b_s|\ldots, \mathbf{Y}) \propto p(a_s|d)\, p(b_s|e) \prod_{k=1}^{S} \prod_{l=1}^{S} p(y_{i,j,t}|A_{k,l,s}, m_{k,l,s})$$

for the marginal predictive $p(y_{i,j,t}|A_{k,l,s}, m_{k,l,s})$ as defined in step (b), $A_{k,l,s} = \{(i,j,t) : i \neq j, \zeta_t = s, \xi_{i,\zeta_t} = k, \xi_{j,\zeta_t} = l\}$ and $m_{k,l,s} = |A_{k,l,s}|$. Since no direct sampler is available for this distribution, we update each pair by using a random-walk Metropolis–Hastings algorithm with bivariate log-normal proposals:

$$(\log(a_s^{(p)}), \log(b_s^{(p)}))^{\mathrm{T}} | (a_s^{(c)}, b_s^{(c)})^{\mathrm{T}} \sim N\{(\log(a_s^{(c)}), \log(b_s^{(c)}))^{\mathrm{T}}, \boldsymbol{\Sigma}_{ab}\}$$

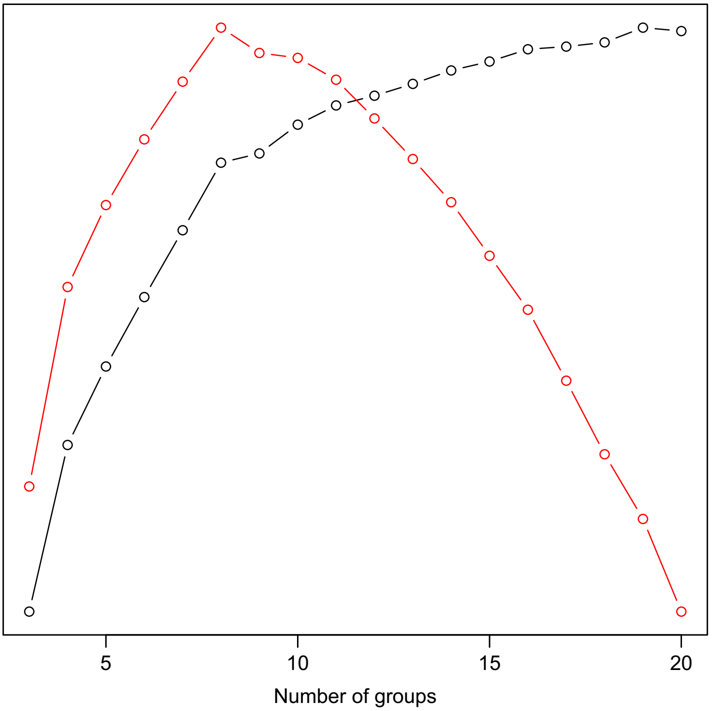


Fig. 10. Integrated classification likelihood criterion (○) for number of groups 3–20 (the optimal number of groups is $Q=8$): log-likelihood

where $\mathbf{\Sigma}_{ab}$ is a tuning parameter matrix chosen independently for diagonal and off-diagonal pairs of parameters.

(f) The parameters of the Poisson–Dirichlet process (α_s, β_s) can be jointly updated by using the algorithm that was described in Escobar and West (1995).

(g) The posterior full conditional distributions for the hyperparameters $d_{\mathrm{O}}, e_{\mathrm{O}}, d_{\mathrm{D}}$ and e_{D} correspond to gamma distributions with shape parameter $cS^* + 1$ and rate parameters $(\Sigma_{S^*} a_{s,\mathrm{O}} + \lambda_d)$, $(\Sigma_{S^*} b_{s,\mathrm{O}} + \lambda_e)$, $(\Sigma_{S^*} a_{s,\mathrm{D}} + \lambda_d)$, $(\Sigma_{S^*} b_{s,\mathrm{D}} + \lambda_e)$ respectively.

Appendix C: Convergence diagnostics

We performed two runs with different initial values and monitored the log-posterior of the model to check convergence. The mixing in these types of model tends to be slow so we thinned the chain every five iterations for a total of 20000 final iterations out of 100000 initial iterations. Fig. 9 displays the trace plots for the two Markov chain Monte Carlo chains, and the results for Gelman and Rubin's R-statistic for convergence diagnostics, where we obtained values close to 1 suggesting that the chains have converged. We utilized the results from the Gaussian emission model in Section 2 to initialize the chains in 'good' starting points to facilitate convergence.

Appendix D: Results for dynsbm

Implementation of the model of Matias and Miele (2017) requires the number of groups to be specified in advance. Fig. 10 displays the integrated classification likelihood for values between 3 and 20 of the number of groups by using `dynsbm`. According to these results the optimal number of groups is 8. We use this number of groups to fit the model and perform the one-step-ahead out-of-sample predictions that were presented in Section 5.2.

References

Adamic, L., Brunetti, C., Harris, J. H. and Kirilenko, A. A. (2010) Trading networks. *Technical Report*. Sloan School of Management, Massachusetts Institute of Technology, Cambridge.
Airoldi, E., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008) Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, **9**, 1981–2014.
Aldous, D. J. (1981) Representations for partially exchangeable arrays of random variables. *J. Multiv. Anal.*, **11**, 581–598.
Ang, A. and Bekaert, G. (2002a) International asset allocation with regime shifts. *Rev. Finan. Stud.*, **15**, 1137–1187.
Ang, A. and Bekaert, G. (2002b) Regime switches in interest rates. *J. Bus. Econ. Statist.*, **20**, 163–182.
Banks, D. and Carley, K. M. (1996) Models for network evolution. *J. Math. Sociol.*, **21**, 173–196.
Barucca, P., Lillo, F., Mazzarisi, P. and Tantari, D. (2017) Detectability thresholds in networks with dynamic link and community structure. *Preprint arXiv:1701.05804*. University of Zurich, Zurich.
Bertoin, J. (2006) *Random Fragmentation and Coagulation Processes*. Cambridge: Cambridge University Press.
Boyd, N., Harris, J. and Nowak, A. (2011) The role of speculators during periods of financial distress. *J. Alterntv. Invest.*, **14**, 10–25.
Cranmer, S. and Desmarais, B. (2011) Inferential network analysis with exponential random graph models. *Polit. Anal.*, **19**, 66–86.
Desmarais, B. and Cranmer, S. (2012) Statistical mechanics of networks: estimation and uncertainty. *Physica* A, **391**, 1865–1876.
Erdös, P. and Rényi, A. (1959) On random graphs. *Publ. Math.*, **6**, 290–297.
Escobar, M. and West, M. (1995) Bayesian density estimation and inference using mixtures. *J. Am. Statist. Ass.*, **90**, 577–588.
Filardo, A. (1994) Business cycle phases and their transitional dynamics. *J. Bus. Econ. Statist.*, **12**, 299–308.
Frank, O. and Strauss, D. (1986) Markov graphs. *J. Am. Statist. Ass.*, **81**, 832–842.
Friel, N., Rastelli, R., Wyse, J. and Raftery, A. E. (2016) Interlocking directorates in Irish companies using a latent space model for bipartite networks. *Proc. Natn. Acad. Sci. USA*, **113**, 6629–6634.
Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**, 457–511.
Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. M. (2009) A survey of statistical network models. *Found. Trends Mach. Learn.*, **2**, 129–233.
Green, P. and Richardson, S. (2001) Modelling heterogeneity with and without the Dirichlet process. *Scand. J. Statist.*, **28**, 355–375.
Guidolin, M. and Timmermann, A. (2005) Economic implications of bull and bear regimes in UK stock and bond returns. *Econ. J.*, **115**, 111–143.
Hanneke, S., Fu, W. and Xing, E. P. (2010) Discrete temporal models of social networks. *Electron. J. Statist.*, **4**, 585–605.
Harris, J., Christie, W. and Schulta, P. (1994) Why did NASDAQ market makers stop avoiding odd eighth quotes? *J. Finan.*, **49**, 1841–1860.
Hatfield, J. W., Kominers, S. D., Nichifor, A., Ostrovsky, M. and Westkamp, A. (2012) Stability and competitive equilibrium in trading networks. *Technical Report*. Stanford University, Stanford.
Heard, N., Weston, D., Platanioti, K. and Hand, D. (2010) Bayesian anomaly detection methods for social networks. *Ann. Appl. Statist.*, **4**, 645–662.
Hoff, P., Raftery, A. and Handcock, M. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.
Holland, P. and Leinhardt, K. (1981) An exponential family of probability distributions for directed graphs. *J. Am. Statist. Ass.*, **76**, 33–65.
Kemp, C., Tenenbaum, J. B., Griffiths, T., Yamada, T. and Ueda, N. (2006) Learning systems of concepts with an infinite relational data. In *Proc. 21st A. Conf. Artificial Intelligence, Boston*, pp. 381–388. American Association for Artificial Intelligence Press.
Kim, C., Morley, J. and Nelson, C. (2001) Does an international tradeoff between risk and return explain mean reversion in stock prices? *J. Empir. Finan.*, **8**, 403–426.
Kolacyzk, E. D. (2009) *Statistical Analysis of Network Models*. New York: Springer.
Lau, J. W. and Green, P. (2007) Bayesian model based clustering procedures. *J. Computnl Graph. Statist.*, **16**, 526–558.
Leifeld, P., Cranmer, S. and Desmarais, B. (2014) xergm: extensions of exponential random graph models. *R Package*. University of Glasgow, Glasgow.
Matias, C. and Miele, V. (2017) Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Statist. Soc.* B, **79**, 1119–1141.
Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Computnl Graph. Statist.*, **9**, 249–265.
Newman, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

Nowicki, K. and Snijders, T. A. B. (2001) Estimation and prediction for stochastic blockstructures. *J. Am. Statist. Ass.*, **96**, 1077–1087.
Ozsoylev, H. N., Walden, J., Yavuz, R. and Bildik, M. D. (2010) Investor networks in the stock market. *Technical Report*. University of California at Berkeley, Berkeley.
Perez-Quiroz, G. and Timmermann, A. (2000) Firm size and cyclical variations in stock returns. *J. Finan.*, **55**, 1229–1262.
Perry, P. O. and Wolfe, P. J. (2013) Point process modelling for directed interaction networks. *J. R. Statist. Soc.* B, **75**, 821–849.
Pitman, J. (1995) Exchangeable and partially exchangeable random partitions. *Probab. Theory Reltd Flds*, **102**, 145–158.
Pitman, J. and Yor, M. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**, 855–900.
Rabiner, L. and Juang, B. (1986) An introduction to hidden Markov models. *IEEE Appl. Signl Process. Mag.*, **3**, 4–15.
Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C. and Samatova, N. F. (2015) Anomaly detection in dynamic networks: a survey. *Computnl Statist.*, **7**, 223–247.
Robert, C. and Casella, G. (2005) *Monte Carlo Statistical Methods*, 2nd edn. New York: Springer.
Robinson, L. and Priebe, C. (2013) Detecting time-dependent structure in network data via a new class of latent process models. *Preprint arXiv:1212.3587v2*. Drexel University, Philadelphia.
Rodriguez, A. (2012) Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statist. Anal. Data Minng*, **5**, 218–234.
Rodriguez, A. and Reyes, P. E. (2013) On the statistical properties of random blockmodels for network data. *Technical Report*. University of California at Santa Cruz, Santa Cruz.
Ryden, T., Terasvirta, T. and Asbrink, S. (1998) Stylized facts of daily return series and the hidden Markov model. *J. Appl. Econmetr.*, **13**, 217–244.
Sarkar, P. and Moore, A. W. (2006) Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems 18* (eds Y. Weiss, B. Schölkopf and J. C. Platt), pp. 1145–1152. Cambridge: MIT Press.
Sewell, D. K. and Chen, Y. (2015) Analysis of the formation of the structure of social networks by using latent space models for ranked dynamic networks. *Appl. Statist.*, **64**, 611–633.
Snijders, T. A. B., van de Bunt, G. G. and Steglich, C. E. G. (2010) Introduction to actor-based models for network dynamics. *Socl Netwrks*, **32**, 44–60.
Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006) Sharing clusters among related groups: hierarchical Dirichlet processes. *J. Am. Statist. Ass.*, **101**, 1566–1581.
Wang, H., Tang, M., Park, Y. and Priebe, C. (2014) Locality statistics for anomaly detection in time series of graphs. *IEEE Trans. Signl Process.*, **62**, 703–717.
Wang, Y. and Wong, G. (1987) Stochastic blockmodels for directed graphs. *J. Am. Statist. Ass.*, **82**, 8–19.
Xing, E. P., Fu, W. and Song, L. (2010) A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Statist.*, **4**, 535–566.
Xu, K. S. (2015) Stochastic block transition models for dynamic networks. In *Proc. 18th Int. Conf. Artificial Intelligence and Statistics*.
Xu, K. S. and Hero, A. O. (2014) Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Selctd Top. Signl Process.*, **8**, 552–562.
Xu, Z., Tresp, V., Yu, K. and Kriegel, H. (2006) Infinite hidden relational models. In *Proc. 21st A. Conf. Artificial Intelligence, Boston*. American Association for Artificial Intelligence Press.
Yang, T., Chi, Y., Zhu, S., Gong, Y. and Jin, R. (2011) Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.*, **82**, 157–189.
Yau, C., Papaspiliopoulos, O., Roberts, G. O. and Holmes, C. (2011) Bayesian non-parametric hidden Markov models with applications in genomics. *J. R. Statist. Soc.* B, **73**, 37–57.
Zaloom, C. (2004) Time, space and technology in financial networks. In *The Network Society: a Cross-cultural Perspective* (ed. M. Castells), pp. 198–216. Cheltenham: Elgar.